Correspondence between a Genetic Network Inferred from Expression Profiles by Strategic Gene Disruptions and the Known Regulatory Relationships

Sachiyo Aburatani[1] **Katsuhisa Horimoto**[2] **Hiroyuki Toh**[3]
sachiyo@grt.kyushu-u.ac.jp horimoto@post.saga-med.ac.jp toh@beri.or.jp

Ayumi Shinohara[4] **Keisuke Iida**[4] **Kosuke Tashiro**[1]
ayumi@i.kyushu-u.ac.jp k-ihda@i.kyushu-u.ac.jp ktashiro@grt.kyushu-u.ac.jp

Satoru Kuhara[1]
kuhara@grt.kyushu-u.ac.jp

[1] Laboratory of Molecular Gene Technics, Graduate School of Genetic Resources Technology, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan

[2] Laboratory of Mathematics, Saga Medical School, Nabeshima 5-1-1, Saga 849-8501, Japan

[3] Department of Bioinformatics, Biomolecular Engineering Research Institute, Furuedai,6-2-3 Suita, Osaka 565-0874, Japan

[4] Department of Informatics, Kyushu University 33, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan

1 Introduction

The recent advance of microarray techniques provides a chance to comprehensively understand the regulatory relationship in a genome scale. In particular, monitoring of gene expression by a gene disruption is one of the strategies to scrutinize the regulatory systems between genes. Here, we present 521 gene profiles monitored in 107 disruptants, and a network that is inferred from the profiles by a newly developed method is consistent with the known regulatory systems.

2 Method and Results

2.1 Expression Profile Data

We monitored the transcriptional levels of 5871 mRNA in 107 disruptants in our laboratory, which contain 64 mutant cells where the transcription factors are deleted. In reference to the knowledge in *Saccharomyces cerevisiae* Proteome Database [3], 94 transcription factors whose regulating genes are clearly identified were found, and the 521 genes in control by 94 factors were selected from 5871 profiles. Thus, we construct a profile set of 521 genes in 107 gene disruption conditions.

2.2 Algorithm for Inferring a Genetic Network

Our system (*ASIAN*: Automatic System for Inferring A Network) is constructed based on two newly developed methods: the automatic determination of cluster number [1] and the inference of a network by graphical Gaussian modeling (GGM) [2]. Thus, *ASIAN* is designed to infer a network only by

inputting the expression profile data. The algorithm is composed of the following three steps: (1) the hierarchical clustering of profiles, (2) the determination of cluster number by the estimation of independence between clusters at each node by the variation inflation factor, and (3) the application of GGM to the profiles that are averaged over the profiles in each clusters obtained in (2). By calculation of the partial correlation coefficient in GGM, it is evaluated whether the clusters are connected or not.

2.3 Inferred Network

ASIAN automatically grouped 507 genes into 37 clusters in which the number of members are ranged from 1 to 81, and among 666 connections between 37 clusters, 294 connections were cut. Consequently, it was inferred that 372 (55.9%) connections between 37 clusters were established (Fig. 1).

2.4 Correspondence Between the Inferred Network and the Known Regulatory Systems

In the 1203 gene pairs in the known regulatory systems, 813 gene pairs were found in the connected or the same clusters. Consequently, 67.6% of gene pairs in the known systems were correctly inferred by *ASIAN*. This correctness is clearly larger than the fraction of the established connections to all connections (55.9%). Furthermore, the consistency between the inferred network and the knowledge of regulatory systems strongly suggests the feasibility of the present analysis tool and the accuracy of profile monitoring experiments.
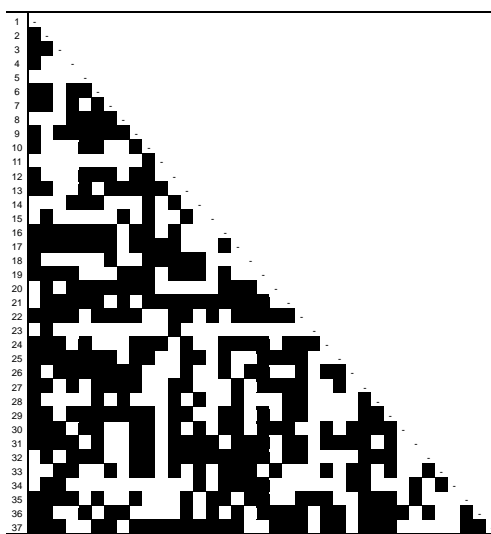

Figure 1: Inferred Network between 37 Clusters by *ASIAN*. The open cell indicates the connection is broken off, and the closed cell indicates that the connection is established.

3 Discussions

Details of the correspondence between the inferred network and the known systems are also investigated. Furthermore, newly discovered relationships between the genes based on the inferred network will be discussed.

References

[1] Horimoto, K. and Toh, H., Statistical estimation of cluster boundaries in gene expression profile data, *Bioinformatics*, in press.

[2] Toh, H. and Horimoto, K., Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling, *Bioinformatics*, in press.

[3] http://www.proteome.com/